DOCUMENT RESUME

ED 195 984 CS 005 805

AUTHOR Davidson, Emma Sue
TITLE An Analysis of Fourth and Sixth Grade Reader Performance Using Cloze Tests, Group Reading Inventories, an Informal Reading Inventory, and a Standardized Reading Test.
PUB DATE Nov 80
NOTE 41p.: Tables may be marginally legible.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cloze Procedure; *Comparative Analysis; Elementary Education; *Informal Reading Inventories; *Reading Ability; *Reading Research; Reading Tests; *Standardized Tests; Test Validity

ABSTRACT

A study was conducted to explore the issue of testing to determine reading levels of students. A group of 624 fourth and sixth grade students from 13 schools participated in the study, which compared results from an informal reading inventory (IRI), a standardized achievement test, a group reading inventory (GRI), and a cloze test. Pupil performance was compared in social studies and in science for the group reading inventory and the cloze test. Correlations between the variables were computed and differences between correlations were tested. To further explore the relations among variables, factor analysis was used. An analysis of variance was performed comparing the two grades on the cloze science and social studies and the GRI science and social studies tests. The analyses showed that the strongest relationships were between the cloze and GRI test results and the weakest between the IRI and the cloze and GRI tests. The standardized test and IRI correlated significantly. In the factor analysis, the standardized test and IRI loaded on the "general reading" factor. Cloze science and social studies and GRI science and social studies loaded on the "content reading and concept processing" factor. The cloze and GRI tests correlated moderately and were found to be more highly correlated with one another than with the standardized test. (Author/FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED195984

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

AN ANALYSIS OF FOURTH AND SIXTH GRADE READER PERFORMANCE USING CLOZE TESTS, GROUP READING INVENTORIES, AN INFORMAL READING INVENTORY, AND A STANDARDIZED READING TEST

by

Dr. Emma Sue Davidson

November 1980

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Emma Sue Davidson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS505805

Abstract

The complex issue of testing to determine reading levels of students was explored. This study compared results from an informal reading inventory (IRI), a standardized test (SRA), a group reading inventory (GRI), and a cloze test when used to measure reading performance in grades 4 and 6. Comparisons were made using test scores and reading levels. Pupil performance was compared in social studies and in science areas for the group reading inventories and the cloze tests for grades 4 and 6. A population of 624 students in grades 4 and 6 from 13 schools in Northeast Arkansas and Southeast Missouri was administered the SRA achievement test, a science and social studies group reading inventory; and 50 students were administered an informal reading inventory (April & May, 1979). Correlations between the variables were computed and differences between correlations were tested by Fisher's z test. To further explore the relations among variables, factor analysis was used. When comparing reading levels on the four reading tests, Kendall's Tau B was used to analyze the data. A significance test on the differences between correlation coefficients was performed. An analysis of variance was performed comparing grades 4 and 6 on the cloze science and cloze social studies and the GRI science and GRI social studies. Specifically, a 2 (grade) x 2 (content: science vs. social studies) ANOVA with repeated measures was employed. When analyzing test scores for grades 4 and 6, the

strongest relationships were between the cloze and GRI test results. The weakest were correlations of the IRI to the cloze and GRI tests. SRA and IRI correlated significantly. In the factor analysis, SRA and IRI loaded on Factor 2, which was labeled "general reading." Cloze science, cloze social studies, GRI science and GRI social studies loaded heavily on Factor 1, which was labeled "content reading and concept processing." Kendall's Tau B was computed to determine the similarity of tests when analyzing reading level placement. For both grades 4 and 6, the intercorrelations were generally low, with the highest relation between the SRA and IRI. The cloze and GRI tests correlated moderately and were found to be more highly correlated with one another than with SRA. An ANOVA was used to determine the effects of grade and subjects. This analysis of variance showed that there was a main effect of grade and that there was interaction between grade and subject. Based on the findings of this study, it was concluded that: (1) the diagnosis of reading ability needs to go beyond the measurement of general reading if the user intends to examine the reader's ability to apply the process of reading content materials in grades 4 and 6; (2) students may be performing poorly in content subjects not because they lack general reading ability, but because they lack the ability to apply the necessary reading skills for content reading; (3) since the cloze procedure has been validated as a useful measurement procedure in reading and the data from this study indicates a strong observed relationship between GRI and cloze test scores, there is an indication of a degree of support for using the GRI in

grades 4 and 6; and (4) results did not support the contention that standardized test results produce an inflated reading score when compared to reader achievement on an IRI.

RATIONALE

Measurement and evaluation in reading is usually concerned with how well a student reads. There is a need to determine the instructional reading level of students in an effort to find suitable instructional materials and methods. The evaluation of pupil reading performance and the interpretation of test results is of paramount importance in the teaching of reading. Exploration for valid assessment measures to determine the instructional level of students leads to an examination of currently used reading tests. What is available to help teachers assess reading levels of students varies from school to school.

A review of research has revealed that (1) standardized testing, even though commonly used, is adversely criticized by many; (2) standardized test scores tend to produce inflated estimates of achievement; therefore, they may not be helpful when used to establish pupils' instructional level; (3) an effort is being made to find alternative tests to replace, or to supplement, the standardized tests; (4) the IRI is acclaimed by many reading specialists to be the best test for use in matching the student with instructional materials; (5) the IRI may not be used extensively enough (because of the time required to administer the test) to be of great value to educators for determining instructional levels of students; (6) the GRI has been

developed to use instead of the IRI because it can be given to a class in 40 minutes; (7) the GRI has not been researched extensively and there is a lack of information concerning the validity of the GRI; (8) no research is evidenced contrary to the fact that the GRI is a good test for determining the reading levels of pupils; (9) cloze testing has been extensively researched, and there is much support for its use for student placement in the appropriate instructional material; and (10) most cloze tests have used narrative materials instead of content materials, and there is a need for studies to use content materials.

This research deals with a combination of tests (standardized test, IRI, GRI, and cloze test) for grades 4 and 6 which have received no comparative research in previous investigations.

Table 1 (page 3) gives the kinds of tests used in this research; the name of the tests; the approved grade level of the test; if the test was administered to a group or administered individually; time required to administer the test; subject area of the test; type of scores rendered; and how scores were classified into categories.

THE STUDY

The Sample

The sample for this investigation was composed of 13 fourth-grade classes and 11 sixth-grade classes from schools in Northeast

Table 1

Information Table Concerning Mensurement Instruments

Kind	Name	Appro. Grade Level	Adm. to	Adm. Time (Min.)	Subj. Area	Score Type	Classify into Categories by Criteria
Stand.	SRA Achiev. Series	4-6	Gr.	48	Gen. Comp. & Voc.	Mean, Raw & GE	Grade 4: Above 5.2 (IND); 4.2-5.2 (INST); Below 4.2 (FRUST). Grade 6: Above 7.2 (IND); 6.2-7.2 (INST); Below 6.2 (FRUST)
IRI	Classroom Reading Inventory	2-10	Ind.	25	Gen. Oral graded paragraphs	Word recog. err. & comp. err. to obtain rdg. lev. (1)IND: (2)INST: (3)FRUST	WR Errors Category: IND 2; INST 5; FRUST 10+. COMP Errors Category: IND 0-1; INST 1½-2; FRUST 2½+
GRI	Content Inventories	4-12	Gr.	40	Sci. & Soc. Stud.	% corr. for obtaining rdg. lev. (1)IND; (2)INST.; (3)FRUST	Table 1-Scoring Group Reading Inventories
Cloze	Content Inventories	4-12	Gr.	40	Sci. & Social Stud.	% corr. for obtaining rdg. lev. (1)IND;(2)INST; (3)FRUST	58 to 100 percent = IND Level - Category 1; 44 to 57 percent = INST Level - Category 2; 43 percent or below = FRUST Level - Category 3

Table 1-Scoring Group Reading Inventories

Number Missed	Number Correct	Number Possible	Percent Correct	Placement Category
0	14	14	100	Independent
1	13	14	93	Level
2	12	14	86	
3	11	14	79	Instructional
4	10	14	71	Level
5	9	14	64	
6 or more	8 or less	14	57	Frustration Level

Arkansas and Southeast Missouri. The schools were within a distance of 150 miles of each other. Each school had more than one fourth and one sixth grade class. One class from grades 4 and 6 in each school was selected for testing. All students in the designated classrooms were included in the study. There were 340 fourth-grade students and 284 sixth-grade students in the study. Since the sample of 624 students was comprised of members of intact classrooms, it seemed reasonable to assume that a representative sample had been obtained and hence, the results of the study could be generalized to most grade 4 and grade 6 students. Furthermore, it was assumed that the intact classrooms gave a racial balance that was approximately the same as would be found in most schools in Northeast Arkansas and Southeast Missouri in grades 4 and 6.

IQ scores could not be obtained for all students. For the 44 percent that was obtained, the range of IQ was 71 to 143, and the mean score was 107.

Measurement Instruments

The following instruments were used in the study:

SRA Achievement Series (Thorpe, Lefever, & Naslund, 1978). Reading achievement was determined from the grade equivalent of the total reading subtest score of the SRA. Students in grade 4 and 6 were administered the SRA Elementary Test. The tests were administered and scored according to the directions in the manual.

Reading tests

According to a review by Fry (which was based on the evaluation of an earlier than 1978 edition of the test), the reading tests of the SRA achievement series are carefully thought out, nicely prepared, and widely standardized. They can be recommended for group testing of average and above-average students. The two reading scores given were vocabulary and comprehension. The average of the vocabulary and comprehension scores gives a total reading score. The supplementary material, such as scoring key and manual for the examiner, are all well done (Buros, 1968, p. 325).

The grade equivalent score (GE) is derived during standardization to answer the following question: How does a student's raw score compare with scores obtained nationally by students in various grades? The GE is defined by the median raw score, the score that falls at the 50th percentile of a national sample of students in a given grade and month in school. For example, a GE of 7-3 represents the median raw score of a national sample of students who took the same test in the third month of grade 7.

Informal Reading Inventory. The informal reading inventory used in this study was the Classroom Reading Inventory (Silvaroli, 1975). Within recent years, a number of informal reading inventories, including Silvaroli's inventory, have been published in formats that tend to make them formal. Rigid steps are outlined for the administration and technical skills required for the interpretations of their results (Zintz, 1977, p. 71).

The Classroom Reading Inventory by Silvaroli is probably the best known of the commercially-developed informal reading inventories. The stories are well written, and they are graded according to accepted readability formulas (Ekwall, 1977, p. 279). Zintz (1977, p. 454) also presents the Classroom Reading Inventory as an exemplary informal reading inventory. This inventory is designed to be used by teachers in grades 2 through 10. It consists of graded word lists, graded oral paragraphs, and graded spelling surveys. It is completely contained in one spiral-bound booklet. The directions are well written and precise enough so that a beginning teacher can use it with little difficulty, yet it is sophisticated enough for use by reading clinicians or school psychologists (Ekwall, 1977, p. 279). This study used the graded oral paragraphs, Form C.

Group Reading Inventory. A group reading inventory is an informal test administered to a group of students to determine if specific reading materials are on the independent, instructional, or frustration level (Kaiser, 1975, pp. 42-43). This study used Content Inventories: English, Social Studies, Science (McWilliams & Rakes, 1979) as the content-related diagnostic materials for the areas involving science and social studies in grades 4 and 6.

The prose passages ranged from 422 to 496 words. The comprehension questions were written using the questioning skills of main idea, context, detail, and inference. Administration and scoring procedures were consistent with directions prescribed in Content

Inventories: English, Social Studies, Science (McWilliams & Rakes, 1979, p. xiv). Questions of a main idea, detail, and contextual type are passage dependent, and were scored accordingly. Responses to the four inferential questions were interpreted with appropriate answers (McWilliams & Rakes, 1979, pp. 84-86, 90-92, 111-113, 117-119, 146, 148).

To administer the GRI, the examiner selected the inventory from the desired content area and at the readers' present grade level in school. Each student was given a copy of the test. The examiner read the following instructions to the class:

> You should read carefully and try to remember as much information as you can. After you have finished reading silently, raise your hand and I will collect your story and return a group of questions about the selection. You are to answer each question based upon what you can remember from the story. Do you have any questions? (McWilliams & Rakes, 1979, p. xiii)

The GRI was administered with a maximum of 40 minutes testing time. After scoring the GRI, the students' scores were placed within one of three reading categories as shown in Table 1.

Cloze procedure. The cloze procedure is a test for measuring the effectiveness of communication from the writer or speaker to the readers or listeners. The term "cloze" is derived from the Gestalt concept of "closure," which is the attempt to complete a structural whole by filling in what is missing. The cloze procedure requires the systematic deletion of every "nth" word in a prose selection, and evaluates how accurately the reader supplies the words that

were deleted (Bormuth, 1968). This study used Content Inventories: English, Social Studies, Science (McWilliams & Rakes, 1979) as the diagnostic source for the areas involving science and social studies in gradcs 4 and 6. The tests for both grades 4 and 6 consisted of approximately 450 word prose sample. The first and last sentences of the selections were left intact. In the other sentences, every seventh word was deleted and a blank of equal length was used to replace each word that was deleted. There were 50 blanks and each correct answer was equivalent to two percentage points. Only exact word replacements were checked correct.

The examiner selected the cloze test from the social studies and science content areas at the students' present grade level in school. Many students were unfamiliar with the cloze procedure; therefore, a practice session using a cloze test was provided prior to the administration of the test to be used in this research. Following the practice session, each student was given a copy of the cloze test. The examiner read the following directions to the class:

> You will notice that several words have been left out of this reading passage. You are to fill in the blanks with a word you think makes sense. Guessing is permitted, so try to fill in as many blanks as possible. Do you have any questions? (McWilliams & Rakes, 1979, p. xi)

The cloze test was administered with a maximum of 40 minutes testing time. When scoring reader responses on the cloze test, only exact word replacements were considered correct. After scoring the test, a percentage correct and placement level was determined by

using a Cloze Placement Table from Content Inventories: English, Social Studies, Science (McWilliams & Rakes, 1979, xii). The criteria for cloze placement level is presented in Table 1, page 3.

Data Collection Techniques

Thirteen fourth-grade classes and 11 sixth-grade classes from schools in Northeast Arkansas and Southeast Missouri were used in the study. The total population of 624 students in the study was administered a cloze test in science and one in social studies, a GRI in science and one in social studies, and a standardized reading test. Informal reading inventories (IRI) were administered to 25 randomly selected students in three fourth-grade classes, and to 25 randomly selected students in three sixth-grade classes, for a total of 50 students who were administered the IRI.

The testing for the study was carried out by the investigator in May of 1979. The cloze test in science and in social studies, and the group reading inventories in science and in social studies were administered to an intact classroom of students by the researcher with the classroom teacher's assistance. The amount of time allotted was 40 minutes per test. One day, the cloze science test and the group reading inventory in science was given, and the following day, the cloze test in social studies and the group reading inventory in social studies was given. All tests were given before 1:00 P.M. The order of test presentation was alternated in order to limit practice effects.

Reading tests

The SRA Achievement Series was administered (April, 1979) by the assigned school examiner as part of a district-wide testing program. An IRI was administered (May, 1979) by the investigator to individual students. Twenty-five minutes per student was the time allotted for the administration of the informal reading inventory; however, in most instances, less time was necessary.

The cloze tests in science and in social studies, and the group reading inventories in science and in social studies were scored by directions that were given in the manual (McWilliams & Rakes, 1979, p. xii). Results were recorded by listing score correct, percent correct, and placement level for each test for each student.

The informal reading inventories that were administered to 25 sixth-grade students and 25 fourth-grade students were scored after each individual test. Scores were recorded of percent correct and placement level for each student as suggested by the test author (Silvaroli, 1975).

The standardized reading score was obtained from the SRA Achievement Series from official school records in each school. Only the subtests of comprehension and vocabulary were obtained. The total reading score was recorded as raw score, grade equivalent, and percentile.

The SRA scores and the IQ score that were recorded for the students in the study were recorded by a school official in each

school. Before the school official returned the information form to the researcher, the names of the students and the name of the school were clipped from the form. This protected the privacy of the individual and the school.

Design for Analysis of Data

Data was analyzed in relation to each hypothesis. For all analysis, the .05 level of significance was used.

Hypothesis 1 compared test scores on the four reading tests (standardized, IRI, GRI, and cloze). A Pearson product-moment correlation was used to analyze the data. The significance of correlations was tested by *t* tests, and differences between correlations were tested by Fisher's *z* test (Glass & Stanley, 1970, pp. 265-268). To further explore the relations among the variables, factor analysis was used.

The factor analysis procedure in this study consisted essentially of (1) calculating intercorrelations among these tests to find those which tend to "load on" or measure the same factors; (2) deducing what the factor is that a particular subgroup of correlated tests are measuring in common; (3) distilling out a satisfactory number of factors; and (4) giving each factor a verbal description (Kerlinger, 1964, pp. 650-684).

Hypothesis 2 compared reading levels on the four reading tests (SRA, IRI, GRI, and cloze). Kendall's Tau B was used to analyze the data (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975,

pp. 288-290). A significant test on the differences between correlation coefficients was performed as described in Ferguson (1976, pp. 183-185).

Hypothesis 3 stated that the relations in Hypothesis 1 and Hypothesis 2 will be evidenced for grades 4 and 6. The correlation between the GRI and cloze test will be lower in grade 4 than in grade 6. The difference between correlation coefficients was tested using Fisher's normal transformation of the correlation coefficient as described by Ferguson (1976, pp. 182, 184, 493).

To further explore the relationship between test results, an analysis of variance was performed comparing grades 4 and 6 on the cloze science and cloze social studies and the GRI science and GRI social studies. Specifically, a 2 (grade) x 2 (content: science vs. social studies) ANOVA with repeated measures was employed. Content was the within-subject variable (repeated measure) (Nie, et al, 1975, pp. 410-415).

Results

Presentation of the results of the data analysis and the evaluation of the hypotheses on the basis of the results is stated. Each of the original hypotheses is presented along with a discussion of the results. After each hypothesis is stated, a brief description of the statistical method employed and the rationale for its use is presented.

Reading tests

Summary statistics for six reading tests administered to students in grades 4 and 6 are presented in Table 2 (page 14). The SRA test was administered in April, 1979, and all other tests were administered in May, 1979. The statistical analysis for the SRA reading score is based upon grade equivalent score (GE). For basis of comparison, the tests were administered to fourth and sixth graders at 4.8 and 6.8 of the school year. They averaged .8 and .7 grade levels above respectively. Results from each of the remaining test results were analyzed on the basis of percent correct. Fourth and sixth grade average scores are also represented in Table 2 by the reading levels (I = Independent; II = Instructional; and III = Frustration; see Table 1 for explanation of how levels were determined). The sixth-grade averages generally indicate reading Level II, except for GRI, whose mean score was consistent with Level III. The fourth-grade averages present marked differences. Note, for example, the cloze science average reading level of I contrasted to the GRI science and GRI social studies average reading levels of III. These inconsistencies will further be addressed in respect to the stated hypotheses of this investigation.

Hypothesis 1. It was hypothesized that test scores on the six reading tests would be positively correlated, but that the IRI, GRI, and cloze would be more highly correlated with one another than they would to the SRA.

Table 2

Test Data Used in Analysis

	Grade 4								Grade 6						
Name of Test	Lev.	Test Date Mo./Yr.	N	Data	Mean	Reading Level Compar.	S.D.		Lev.	Test Date Mo./Yr.	N	Data	Mean	Reading Level Compar.	S.D.
SRA Achievement Composite Voc. & Comp.	E	4-79	335	Grade Equiv.	5.62	II	1.83		F	4-79	261	Grade Equiv.	7.51	II	2.06
Cloze Science	4	5-79	335	Percent Correct	59.32	I	16.70		6	5-79	241	Percent Correc.	46.89	II	21.98
Cloze S. S.	4	5-79	258	Percent Correct	40.98	III	19.98		6	5-79	231	Percent Correct	53.98	II	20.51
CRI Sci.	4	5-79	334	Percent Correct	47.83	III	23.06		6	5-79	242	Percent Correct	56.27	III	20.90
CRI S. S.	4	5-79	180	Percent Correct	49.97	III	20.89		6	5-79	205	Percent Correct	57.77	III	24.04
IRI	4	5-79	25	Percent Correct	98.00	I	5.00		6	5-79	25	Percent Correct	89.60	II	11.54

Intercorrelations of the reading tests are presented in Table 3 (page 16). Correlations are presented above the diagonal. The number of pairs of scores upon which the correlations are determined are presented below the diagonal. Significant levels are below the number.

As can be seen from Table 3, all significant correlations for grade 4 were positive. The negative correlations between the IRI, the cloze and GRI tests were not significantly different from zero. In general, for grade 4, the strongest relationships were between the cloze and GRI tests. The weakest were those involving the IRI. Clearly, the SRA was not associated with the lowest correlations as had been hypothesized. However, cloze science and social studies and GRI science and social studies test scores still appear to be more highly correlated with one another than with SRA test scores. The IRI test scores, when correlated with the cloze science and social studies and GRI science and social studies tests, tended to show lower relationship with these tests than did the SRA.

In general, the grade 6 results were similar to those found in grade 4; however, the grade 6 results did not produce as explicit differentiations as evidenced in grade 4. Of note in the grade 6 results is the marked higher relationship of the SRA and IRI scores than was found in grade 4. Consistent with the grade 4 results are the higher cloze science and social studies correlations with one

Table 3

Intercorrelations of Reading Test Scores

Name of Test	Grade 4 SRA	Cl. Sci.	Cl. S.S.	CRI Sci.	CRI S.S.	IRI	Grade 6 SRA	Cl. Sci.	Cl. S.S.	CRI Sci.	CRI S.S.	IRI
SRA		.485**	.291**	.506**	.312**	.412*		.191*	.325**	.448**	.469**	.683*
Cl. Sci.	N=330 S=.001	-	.739**	.656**	.665**	-.088	N=221 S=.002	-	.744**	.462**	.569**	-.248
Cl. S.S.	N=255 S=.001	N=254 S=.001	-	.583**	.626**	-.244	N=213 S=.001	N=191 S=.001	-	.466**	.568**	-.128
CRI Sci.	N=329 S=.001	N=331 S=.001	N=254 S=.001	-	.604**	-.296	N=221 S=.001	N=235 S=.001	N=191 S=.001	-	.570**	.102
CRI S.S.	N=177 S=.001	N=176 S=.001	N=178 S=.001	N=175 S=.001	-	-.142	N=187 S=.001	N=164 S=.001	N=200 S=.001	N=165 S=.001	-	-.077
IRI	N=25 S=.020	N=25 S=.337	N=25 S=.119	N=23 S=.085	N=24 S=.253	-	N=25 S=.001	N=24 S=.120	N=25 S=.271	N=24 S=.316	N=25 S=.356	-

*p < .05

**p < .01

Correlations are presented above the diagonal. The number of pairs of scores upon which the correlations are determined are presented below the diagonal. The significant levels are presented below the N.

another than with SRA or IRI. In summary, the SRA correlations with the GRI and cloze science and social studies tests indicate stronger relationship than did the corresponding IRI test correlations. This tendency is contrary to the hypothesized pattern.

To further explore the relations among the variables, factor analysis was used. The factor analysis procedure consisted of (1) calculating intercorrelations among the reading tests, as just reviewed, to find those which tend to "load" or measure the same factor; (2) deducing what factor a particular subgroup of correlated tests were measuring in common; (3) identifying the number of factors; and (4) giving each factor a verbal description. Thus, the objective was to determine how many factors there were, what tests were loading on what factors, and the magnitude of the test loadings.

In treating the statistical results from the different tests, the correlations were computed and a correlation matrix was set up (Table 3). The intercorrelates were necessary to utilize the principal factor method to produce a set of factors. Each factor solution gave the correlation coefficient between each variable and each factor. The rotational method used was the varimax, which is an orthogonal factoring solution. The pairwise procedure, which involves all cases in the computation of all correlation coefficients for which there is complete data, was used.

Table 4 (page 19), which is the terminal solution of the orthogonally rotated factors, presents Factors 1 and 2 for grades 4 and 6. The relationship or loading of each test on the identified factor is shown. For both grades 4 and 6, cloze science, cloze social studies, GRI science, and GRI social studies loaded on Factor 1. IRI showed a negative loading on Factor 1 in both grades 4 and 6. Loading on Factor 2 was SRA and IRI; however, SRA also loaded on Factor 1, but to a lesser extent. Cloze social studies had a negative loading on Factor 2 for both grades.

From inspection of the test items for each test, it appears that the SRA and IRI are concerned with the student's general reading abilities. General reading is composed of a general educational factor plus verbal ability. General educational factors for reading include word recognition and word meaning, grasping details, perceiving relationship, drawing conclusions, and processing materials.

As just described, both tests, SRA and IRI, loaded heavily on Factor 2; thus, it would seem reasonable to label this factor as "general reading." The SRA test, however, required more application of the material read, which could have caused it to load some on Factor 1.

Content material in the area of science and social studies was used in the cloze and GRI tests. The type of material in these tests and the application requirements for testees, clearly shows

Table 4

Varimax Rotated Factor Matrix

Variable	Grade 4		Grade 6	
	Factor 1	Factor 2	Factor 1	Factor 2
SRA	.490	.708	.411	.911
Cloze Science	.869	.122	.839	-.157
Cloze Social Studies	.800	-.095	.811	-.014
GRI Science	.801	.005	.618	.241
GRI Social Studies	.750	-.008	.754	.141
IRI	-.245	.753	-.172	.821
Eigenvalue	2.92059	1.07251	2.62758	1.48997
Percent of Variance	73.1	26.9	63.8	36.2
	73.1	100.0	63.8	100.0

that these tests are measuring something different from "general reading." As described, cloze science, cloze social studies, GRI science, and GRI social studies loaded on Factor 1. It would seem rational to label Factor 1 as "content reading and concept processing."

Hypothesis 1 was, therefore, partially accepted. For both grades 4 and 6, all significant correlations were positive. In general, the SRA correlated markedly stronger with the cloze and GRI tests than did the IRI test scores with the cloze and GRI tests. Cloze and GRI tests were found to be more highly correlated with one another than with SRA. IRI, not the SRA, had the weakest relationship with the other reading tests. Factor analysis showed that for both grades 4 and 6, cloze science, cloze social studies, GRI science, and GRI social studies loaded on Factor 1, which was labeled "content reading and concept processing." The IRI and SRA loaded heavily on Factor 2, which is labeled "general reading." The implication is that the cloze and GRI tests measure one type of reading skill, while the IRI and SRA measure a different type of reading skill.

Hypothesis 2. Reading levels on the six reading tests were hypothesized to be positively correlated with one another; however, IRI, GRI, and cloze were expected to be more highly correlated with one another than they would to the SRA.

Kendall's Tau B, which is a nonparametric correlation, was computed to determine the similarity of the tests being studied. The use of rankings is required for the computation of the coefficients so the rankings of 1 (Independent), 2 (Instructional), and 3 (Frustration) were read in. The student's performance on the IRI, GRI, and cloze test were analyzed by using percent correct to provide a classification (as stated in the test manual) of whether the material that was written on grade level difficulty was on the student's independent, instructional, or frustration level. The SRA used grade equivalent (GE) as indicated in Table 1.

Table 5 (page 22) presents reading level classifications for grades 4 and 6 as determined for each test in the research. The number of students whose performance placed them in specified categories is given. Under the category number is the percentage of students in each classification. For grade 4, IRI, cloze science, and SRA had the highest percentage in the independent level and lowest percentage in the frustration level. GRI social studies, GRI science, and cloze social studies had the highest percentage in the frustration level. SRA, IRI, and cloze social studies classified higher proportions of grade 4 participants into the independent level than the frustration; for cloze science, GRI social studies and GRI science, just the opposite was observed.

Grade 4 results were similar in this comparison of reading level classifications, except for the cloze social studies and cloze science results, where just the reverse was observed.

Table 5

Reading Classification Levels as Determined by Tests for Grades 4 and 6

Grade 4

Category Level		SRA Total Rdg.	Cloze Sci.	Cloze S. S.	GRI Sci.	GRI S. S.	IRI
1. Independent	N	184	195	36	24	11	21
	Percent	(54.9)	(58.2)	(14.0)	(7.2)	(6.1)	(84.0)
2. Instructional	N	81	109	101	87	47	3
	Percent	(24.2)	(32.5)	(39.3)	(26.0)	(26.1)	(12.0)
3. Frustration	N	70	31	120	223	122	1
	Percent	(20.9)	(9.3)	(46.7)	(66.8)	(67.8)	(4.0)
TOTAL	N	335	335	257	334	180	25
	Percent	(100.0)	(100.0)	(100.0)	(100.0)	(100.0)	(100.0)

Grade 6

Category Level		SRA Total Rdg.	Cloze Sci.	Cloze S. S.	GRI Sci.	GRI S. S.	IRI
1. Independent	N	137	47	79	22	39	10
	Percent	(52.5)	(19.5)	(34.2)	(9.1)	(19.0)	(40.0)
2. Instructional	N	52	111	115	96	65	9
	Percent	(19.9)	(46.1)	(49.8)	(39.7)	(31.7)	(36.0)
3. Frustration	N	72	83	37	124	101	6
	Percent	(27.6)	(34.4)	(16.0)	(51.2)	(49.3)	(24.0)
TOTAL	N	261	241	231	242	205	25
	Percent	(100.0)	(100.0)	(100.0)	(100.0)	(100.0)	(100.0)

Table 6 (page 24) presents a summary of the results of Kendall's Tau b analyses for grades 4 and 6. For grade 4, the intercorrelations were generally low, with the highest, r = .731, obtained for the relation between SRA and IRI. A significance test on the differences between correlations coefficients was performed, as described in Ferguson (1976, pp. 183-185). The results make it clear that SRA correlations were comparable to (and sometimes higher than) those for the remaining tests. An unexpected finding was the negative correlations between IRI and GRI science, GRI social studies, and cloze social studies. The GRI and cloze tests correlated moderately and were more highly correlated with one another than with SRA.

As was noted for grade 4, the intercorrelations for grade 6 were generally low, with the highest, r = .674, obtained for the relationship between SRA and IRI. SRA correlations for the remaining tests were low. There were nonsignificant negative correlations between IRI and GRI social studies, cloze social studies, and cloze science. The GRI and cloze tests correlated moderately and were more highly correlated with one another than with SRA.

Hypothesis 2 was partially accepted. All correlations were not positive and significant. The IRI had a negative correlation with all tests except the SRA and cloze science (grade 4) and GRI science (grade 6). The SRA correlated markedly stronger with the

Table 6

Kendall's Tau B Analyses

Grade 4

Variable	Variable	Kendall's Tau B	Significance
SRA	Cloze Science	.332	S = .001
SRA	GRI Science	.377	S = .001
SRA	Cloze Social Studies	.300	S = .001
SRA	GRI Social Studies	.171	S = .007
SRA	IRI	.731	S = .001
IRI	Cloze Science	.179	S = .184
IRI	GRI Science	-.122	S = .179
IRI	Cloze Social Studies	-.235	S = .036
IRI	GRI Social Studies	-.163	S = .104
Cloze Science	GRI Science	.306	S = .001
Cloze Science	Cloze Social Studies	.220	S = .001
Cloze Science	GRI Social Studies	.241	S = .001
Cloze S. S.	GRI Social Studies	.507	S = .001
GRI Science	Cloze Social Studies	.444	S = .001
GRI Science	GRI Social Studies	.473	S = .001

Grade 6

SRA	Cloze Science	.240	S = .000
SRA	GRI Science	.304	S = .001
SRA	Cloze Social Studies	.262	S = .001
SRA	GRI Social Studies	.226	S = .001
SRA	IRI	.674	S = .001
IRI	Cloze Science	-.274	S = .072
IRI	GRI Science	.000	S = .500
IRI	Cloze Social Studies	-.172	S = .182
IRI	GRI Social Studies	-.141	S = .221
Cloze Science	GRI Science	.44[illegible]	S = .001
Cloze Science	Cloze Social Studies	.3[illegible]	S = .001
Cloze Science	GRI Social Studies	.429	S = .001
Cloze S. S.	GRI Social Studies	.495	S = .001
GRI Science	Cloze Social Studies	.437	S = .001
GRI Science	GRI Social Studies	.446	S = .001

cloze and GRI tests than did the IRI. However, cloze and GRI tests were found to be more highly correlated with one another than with the SRA.

Hypothesis 3. It was hypothesized that the relations in Hypothesis 1 and Hypothesis 2 would be evidenced for grades 4 and 6; however, the correlation between GRI tests and cloze tests would be lower in grade 4 than in grade 6.

It can be noted in Table 3 that the correlation between GRI science and cloze science test scores was .656 at fourth-grade level, and .462 at sixth-grade level. The difference between correlation coefficients was tested using Fisher's normal transformation of the correlation coefficient as described in Ferguson (1976, pp. 182, 184). Statistically, the fourth-grade correlation was significantly higher than the sixth-grade relationship, $z = 3.34$, $p < .01$. Thus, on the science tests, the hypothesis is rejected: the fourth-grade and sixth-grade relationship was the reverse of what was anticipated.

The correlation between GRI social studies and cloze social studies test scores was .626 at fourth-grade level, and .568 at sixth-grade level (Table 3). Statistically, the fourth and sixth grade correlations are not significantly different ($Z = .866$, $p > .05$). Therefore, the hypothesis that a lower correlation would be found for grade 6 was not supported.

To further explore the relationship between test results, an analysis of variance was performed comparing grades 4 and 6 on the cloze science and cloze social studies tests. Specifically, a 2 (grade) x 2 (content: science vs. social studies) ANOVA with repeated measures was employed on each. Content was the within-subjects variable (repeated measure).

As shown in Table 7 (page 27), there was a significant main effect of content $F(1,443) = 111.88$, $p < .001$, on cloze test scores. There also was significant interaction between grade and content, $F(1,443) = 380.77$, $p < .001$. There was no main effect of grades, $F(1,443) = .443$.

Follow-up analyses of the interaction showed that for grade 4, social studies and science scores differ significantly. Science scores were significantly higher than social studies scores. For grade 6, however, social studies scores were significantly higher than science scores.

GRI test results for grades 4 and 6 were analyzed using the same ANOVA design as above. As shown in Table 8 (page 28), there was a main effect of grades, $F(1,338) = 22.620$, $p < .01$. There was no main effect of content. Furthermore, there was no significant interaction between grade and content. The grades main effect showed that grade 6 scored higher than grade 4 on GRI science and social studies tests.

Table 7

Analysis of Variance for Cloze Science and Cloze Social Studies

Grades 4 and 6

Source of Variation	df	Sum of Squares	Mean-Square	F	Prob. F Exceeded
Grade Main Effect	1	289.437	289.437	.433	.510
Error	443	295477.00000	666.990		.510
Test Content	1	11404.949	11404.949	111.884*	.001
Grade x Content	1	38814.339	38814.339	380.773*	.001
Error	443	45157.410	101.935		

*Significant at .01

Cell Means for CRI Science and CRI Social Studies

Grades 4 and 6

	Grade 4	Grade 6	Marginal
Cloze S.S.	41.007	55.502	47.229
Cloze Science	61.582	49.192	56.350
Marginal	51.295	52.447	
Count	254	191	

	Standard Deviation	
	Grade 4	Grade 6
Cloze Social Studies	20.080	20.798
Cloze Science	16.418	21.549

Reading tests

Table 8

Analysis of Variance for GRI Science and GRI Social Studies

Grades 4 and 6

Source of Variation	df	Sum of Squares	Mean-Square	F	Prob. F Exceeded
Grade Main Effect	1	17483.25000	17487.25000	22.620*	.001
Error	338	261234.0000	772.88159		
Test Content	1	1.062	1.062	.005	.943
Grade x Content	1	8.43750	8.43750	.041	.839
Error	338	68972.062	204.059		

*Significant at .05

Cell Means for GRI Science and GRI Social Studies

Grades 4 and 6

	Grade 4	Grade 6	Marginal
GRI S.S.	49.725	59.648	54.541
GRI Science	49.582	59.951	54.614
Marginal	49.654	59.799	
Count	175	165	

	Standard Deviation	
	Grade 4	Grade 6
GRI Social Studies	20.941	24.337
GRI Science	22.611	20.332

Hypothesis 3 was rejected in both the science and social studies contrasts with the cloze and GRI scores. In fact, the correlation between the GRI and cloze science scores was significantly higher for fourth graders than that observed for sixth graders.

DISCUSSION

From inspection of the test items appearing on each test, and from the type of reading process that was required of the reader, it is understandable why the SRA test and the IRI loaded on one factor and the GRI and cloze tests loaded on a different factor.

The IRI and SRA tests were composed of general reading materials as opposed to content-related materials. In taking the IRI, the student read (narrative material) orally while the examiner recorded word recognition errors. The examiner asked comprehension questions and recorded the pupil's oral responses. As outlined in the administration manual (Silvaroli, 1975), the word recognition percent of errors and the comprehension percent of errors provided the criteria for placing the student on an independent, instructional, or frustration level. It should be noted that the reading process required on the IRI was of an oral nature.

The SRA loaded on the same factor as the IRI, which was Factor 2 that was labeled "general reading." Performance on the SRA required silent reading of general reading material written in a narrative style. Responses to comprehension questions were multiple choice. The reader response required more application of the material read than did the IPI, which could have caused the SRA to load some on the other factor. Standardized tests have been labeled as general reading achievement tests by reading authorities. Farr (1969) stated that there are valid uses for group tests in the instructional reading program. Farr also said: "First of all, the tests are reliable for comparing students in terms of general reading achievement" (Farr, 1969, p. 85).

The cloze tests and CRI tests revealed variance from some source other than those in the SRA and IRI, since they loaded on a different factor (Factor 1), labeled "content reading and processing." The foundation for Factor 1 (content reading and processing) was general reading; however, this factor extends beyond the general reading factor and requires the reading of content material plus the processing of concepts and principles dealing with the content materials. In 1944, Artley stated that while some relationship exists between tests of general comprehension and tests of comprehension in social studies, there is a high degree of specificity in the factors relating to reading comprehension in social studies (Artley, 1944, p. 37).

Several studies concluded that comprehension of reading materials in subject areas is different in each subject (Artley, 1944; Shores, 1960; Maney, 1958; & Halfter & Douglass, 1960). This belief indicates diagnosis of a student's reading performance in a content area must be extended beyond the general reading ability. "The diagnosis of reading ability, therefore, needs to go beyond an evaluation of general reading power and should examine the reader's ability to apply his reading skills" (Farr, 1969, p. 117). This study found that when using the group reading inventory (GRI) in the subject areas of science and social studies for grades 4 and 6 there was no main effect of content.

In this research, cloze science and cloze social studies were group tests that used content materials. Every seventh word was deleted except in the first and last sentences, which were left intact. The ability to make closures was related to the student's general reading ability plus his ability to process his understanding of the specialized vocabulary. The cloze test loaded high on Factor 1, "content reading and concept processing."

The GRI science and GRI social studies tests were also group tests that used content materials. Like the cloze tests, these tests were based on general reading ability and extended to the ability to process specialized vocabulary and knowledge of the content areas of science and social studies. Performance on the GRI tests required the student to read (silently) the content

materials in the written passage, and respond in writing to 14 questions provided on a response form. This retention and retrieval process of concepts was anticipated to be difficult for readers in both grades 4 and 6, but it was hypothesized to be more difficult for readers in grade 4 than in grade 6.

The cloze tests and GRI tests loaded on Factor 1. This is apparently what would be expected of tests of this kind; that is, diagnosis of a student's reading performance in a content area must extend beyond his general reading ability.

An unexpected finding was the non-significant negative correlation between the IRI and cloze tests and IRI and GRI tests. This relationship could have been due to several causes which will be discussed. The low relationship between these tests could have been due to the fact that a small number of students were used in the IRI sample. It was possible that the IRI sample was not representative of the characteristics in the population.

Differences resulting from comparing oral reading (i.e., IRI) and silent reading (i.e., SRA, cloze, and GRI) performance could have contributed to a lesser relationship of the IRI to the cloze test and GRI gests. As reported by McCracken (1962, p. 369), IRI (oral reading performance) and standardized tests (silent reading performance) do not generate equivalent reading level placement.

The difference in the type of materials used in writing the tests could have caused the low relationship between the IRI and the cloze and CRI tests. An IRI written with content materials and used with fourth and sixth grade students would possibly correlate differently with cloze and CRI tests. Farr stated that reading performance on one informal reading inventory based on one set of materials or set of basal readers could in all likelihood differ from performance on another IRI based on a different set of materials (Farr, 1969, p. 109).

A final possible reason for the non-significant negative correlation between the IRI and content reading tests could be attributed to the scoring criteria as suggested in this IRI. There is wide disagreement between authors of IRI's as to the appropriate scoring procedure (Ekwall, Solis, & Solis, 1973; Kender, 1968; Pikulski, 1974; and Sipay, 1964). On this point, Kender concluded: "It is difficult to recommend any particular criteria for scoring an informal reading test because no one set of criteria has been sufficiently validated" (Kender, 1968, p. 341).

CONCLUSIONS

The following conclusions were based on the findings of this study: (1) the diagnosis of reading ability needs to go beyond the measurement of general reading if the user intends to

examine the reader's ability to apply the process of reading content material in grades 4 and 6; (2) students may be performing poorly in content subjects not because they lack general reading ability, but because they lack the ability to apply the necessary reading skills for content reading; (3) since the cloze procedure has been validated as a useful measurement procedure in reading and the data from this study indicates a strong observed relationship between GRI and cloze test scores, there is an indication of a degree of support for using the GRI in grades 4 and 6; and (4) results did not support the contention that standardized test results produce an inflated reading score when compared to reader achievement on an IRI.

The complex issue of testing to determine reading levels of students has been explored. Continued investigation of this area may be fruitful in giving insights into the aspects of testing for teaching.

Reference List

Artley, A. S. A study of certain relationships existing between general reading comprehension and reading comprehension in a specific subject matter area. Journal of Educational Research, 1944, 37, 464-473.

Buros, O. K. Reading Tests and Reviews, edited by O. K. Buros. Highland Park, New Jersey: Gryphon Press, 1968, 118.

Ekwall, E. E. Diagnosis and Remediation of the Disabled Reader. Boston: Allyn and Bacon, Inc., 1977, 286-290.

_____, Solis, J., & Solis, E. Investigating informal reading inventory scoring criteria. Elementary English, February, 1973, 50, 271-274.

Farr, R. C. Reading: What Can Be Measured? Newark, Delaware: International Reading Association Research Foundation, 1969, 85, 109, 117.

Ferguson, G. A. Statistical Analysis in Psychology and Education. New York: McGraw-Hill, 1976.

Glass, G. V., & Stanley, J. C. Statistical Method in Education and Psychology. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.

Halfter, I. T., & Douglass, F. M. Measurement of college level reading competence in a content area. Journal of Educational Research, 1960, 53, 223, 230.

Kaiser, R. The group diagnostic inventory. Excerpts from Decision Making: An Imperative for Language Learning. Proceedings of the Seventh Annual Language Arts Conference, Memphis State University, January, 1975.

Kender, J. O. How useful are informal reading tests? The Journal of Reading, February, 1968, 115, 337-341.

Kerlinger, F. Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, Inc., 1964.

Maney, E. S. Literal and critical reading in science. Journal of Experimental Education, 1958, 27, 57-64.

McCracken, P. Standardized tests and informal reading inventories. Education, February, 1962, 366-369.

McWilliams, L., & Rakes, T. Content Inventories for English, Social Studies, Science. Dubuque: Kendall/Hunt Publishing Company, 1979.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. Statistical Package for the Social Sciences. New York: McGraw-Hill Publishers, 1975.

Pikulski, J. J. A critical review: informal reading inventories. Reading Teacher, April, 1977, 30, 767-770.

Shores, H. J. Reading science materials for two distinct purposes. Elementary English, 1960, 37, 546-553.

Silvaroli, N. J. Classroom Reading Inventory. Third Edition. Dubuque, Iowa: William Brown Company, Publishers, 1975.

Sipay, E. R. A comparison of standardized reading tests and informal reading inventories. Reading Teacher, January, 1964, 17, 265-268.

Thorpe, L. P., Lefever, D. W., & Naslund, R. A. SRA Achievement Series (Multilevel Edition). Chicago: Science Research Associates, 1978.

Zintz, M. Corrective Reading. Dubuque: William C. Brown Company, 1977, 71-77.